

Citation for published version:

Dorus, S, Wilkin, EC & Karr, TL 2011, 'Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of Drosophila sperm', *BMC Genomics*, vol. 12, 177.
<https://doi.org/10.1186/1471-2164-12-177>

DOI:

[10.1186/1471-2164-12-177](https://doi.org/10.1186/1471-2164-12-177)

Publication date:

2011

[Link to publication](#)

© 2011 Dorus et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Open Access

Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of *Drosophila* sperm

Steve Dorus¹, Elaine C Wilkin¹ and Timothy L Karr^{2*}

Abstract

Background: The evolutionary diversification of gene families through gene creation (and loss) is a dynamic process believed to be critical to the evolution of functional novelty. Previous identification of a closely related family of eight annotated metalloprotease genes of the M17 Merops family in the *Drosophila* sperm proteome (termed, Sperm-LeucylAminoPeptidases, S-LAPs 1-8) led us to hypothesize that this gene family may have experienced such a diversification during insect evolution.

Results: To assess putative functional activities of S-LAPs, we (i) demonstrated that all S-LAPs are specifically expressed in the testis, (ii) confirmed their presence in sperm by two-dimensional gel electrophoresis and mass spectrometry, (iii) determined that they represent a major portion of the total protein in sperm and (iv) identified aminopeptidase enzymatic activity in sperm extracts using LAP-specific substrates. Functionally significant divergence at the canonical M17 active site indicates that the largest phylogenetic group of S-LAPs lost catalytic activity and likely acquired novel, as yet undetermined, functions in sperm prior to the expansion of the gene family.

Conclusions: Comparative genomic and phylogenetic analyses revealed the dramatic expansion of the S-LAP gene family during *Drosophila* evolution and copy number heterogeneity in the genomes of related insects. This finding, in conjunction with the loss of catalytic activity and potential neofunctionalization amongst some family members, extends empirical support for pervasive "revolving door" turnover in the evolution of reproductive gene family composition and function.

Keywords: sperm proteomics, gene duplication, gene family, protease, spermatogenesis, testis

Background

Gene duplication is considered a fundamental process in the generation of genes with novel functions and ultimately the evolution of biological diversity [1]. A wide variety of evolutionary and comparative genomic analyses have been used to infer the origins and fates of newly created genes [2-5]. While more limited in scope, molecular genetic and functional studies [6-9] have provided new insights into the role gene duplication has played in generating novel functions. In *Drosophila*, comparative genomic analyses have revealed an

enrichment of rapidly evolving or lineage specific gene families associated with reproduction [10]. While the functional significance of this remains to be fully determined, the trend can be explained, in part, by creation of male-biased genes through retrotransposition. Recent studies have demonstrated that retrogenes in both *Drosophila* [11,12] and primates [13-16] tend to acquire testis-specific expression. However, we still have limited understanding of why new genes frequently acquire testis expression and whether testis-specific expression is obtained prior to, or following, acquisition of functionality. Nonetheless, several studies suggest that newly generated genes and gene families have had a functional impact on spermatogenesis and on the evolution of sperm [8,12,17].

* Correspondence: tkarr@asu.edu

²Centers for Evolutionary Medicine and Informatics and Infectious Diseases and Vaccinology, Biodesign Institute, Arizona State University, Tempe, AZ 85287-5001, USA

Full list of author information is available at the end of the article

Although new testis-specific genes might be expected to have gained functions in spermatogenesis, there are a limited number of characterized examples of such genes in *Drosophila*. One example is the strict paternal-effect gene *ms(3)K81* (termed “*K81*”), a gene created by retrotransposition prior to the divergence of the melanogaster subgroup [8]. In wild-type eggs fertilized by sperm from K81 males, paternal chromosomes systematically fail to properly separate sister chromatids during the first zygotic division leading to lethality early in embryogenesis [8]. Interestingly, *K81* is expressed only in primary spermatocytes where it presumably regulates aspects of spermatogenesis required for proper sperm function in the egg. A second example is *mojoless* (*mjl*), created approximately 50mya through the retrotransposition of *shaggy* (*sgg*), a glycogen synthase kinase-3 encoding gene [9]. RNAi knockdown of *mjl* resulted in loss of the male germ line and male sterility. Although not yet ascribed a characterized function in the testis or sperm, a third gene, *Sdic*, is a newly created gene encoding a protein present in the sperm tail [17]. *Sdic* is an unusual case of an X-linked chimeric gene specific to *Drosophila melanogaster* that was created through the duplication of *annexinX* (*AnnX*) and subsequent fusion with *Cdic*, a cytoplasmic dynein gene.

The application of mass spectrometry (MS) to the study of sperm has provided our first insights into the dynamic role new gene creation has played in shaping the constituents of the sperm proteome [18] and has raised intriguing questions about how these new genes have impacted the molecular evolution of sperm. For example, analysis of the *Drosophila* sperm proteome revealed 3 novel sperm genes specific to *D. melanogaster* and a further 4 sperm genes created through retrotransposition during *Drosophila* evolution [12]. Amongst these is the retrogene *CG13340*, which encodes S-LAP 7, examined in this study. It is noteworthy that these new sperm components were found to be proteins across diverse functional classes, and both rapidly evolving (in the case of protamine genes, *Mst35a* and *Mst35b*) and highly conserved (X-linked *Tektin* gene cluster).

Many proteolysis-related genes (including peptidases, proteases and inhibitors of proteolysis) have reproductive functions across diverse taxa, including insects [19-21] and mammals [22]. Amongst these, proteolysis-related genes expressed in the *Drosophila* accessory gland have been particularly well studied in terms of their mediation of a wide range of effects on females (see [23] for a recent review). Proteins with proteolytic activity are also found in the *Drosophila* female reproductive tract and a subset of these are encoded by recently created genes [24]. In both mammals and *Drosophila*, the genes involved in reproductive proteolytic

pathways or their targets have been demonstrated to evolve rapidly [22,25,26] and it is hypothesized that this is due to coevolutionary forces associated with sexual conflict (either between males and females or males and males) [27,28]. These observations have been extended to the sperm proteome where an intensified signature of positive selection was observed for membrane and acrosomal sperm proteins that included a diverse set of metalloproteases and protease inhibitors [29].

Although well documented in reproductive tissues generally, only recently have proteolytic and related genes been shown to be present in spermatozoa [18,29-31]. The functional significance and role in fertilization of these classes of proteins is unknown, but demonstration of their presence and activity in sperm is a necessary step towards more targeted studies. Here we present a detailed evolutionary and functional characterization of a family of eight *Drosophila* sperm leucyl aminopeptidases that we have termed Sperm-LAPs (S-LAP 1-8). Computational annotation places the S-LAPs in the M17 family of leucyl aminopeptidases [32]. Of the 10 annotated leucyl aminopeptidases in the *Drosophila* genome, the S-LAPs are specifically expressed in the testis and all encode proteins incorporated in mature sperm [18,33]. Here we describe results of detailed comparative genomic analyses that demonstrate an expansion and diversification of the S-LAP gene family during *Drosophila* evolution. We also provide (i) an independent proteomic analysis confirming the presence of S-LAPs in sperm, (ii) a biochemical analysis of S-LAPs quantity and abundance in sperm, (iii) data indicating that S-LAPs are expressed specifically in the testis and (iv) evidence confirming S-LAP enzymatic activity supporting their functional annotation. This striking specificity of expression and cellular location suggests an unknown functional requirement for aminopeptidases in sperm. Furthermore, evolutionary diversification at the active site indicates that some S-LAPs have lost enzymatic activity and that the putative neofunctionalization of these non-enzymatic S-LAPs may be related to the subsequent expansion of the gene family. These findings thus support a scenario where functional diversification within a gene family may promote further evolutionary changes in gene family composition through the selective retention of newly created genes.

Results

Comparative Genomics of the S-LAP family in *Drosophila*
Gene Ontology (GO) analysis [34] previously identified eight related leucylaminopeptidases of the M17 Merops family with an overall average amino acid identity of 53.7% [18]. We examined their evolutionary relationships by first determining that the annotated aminopeptidase, *granny-smith*, was the closest related paralog to

the S-LAP gene family (17.4% average amino acid conservation with the S-LAP family). *granny-smith* has not been identified in *Drosophila* sperm [33] and has no reported function in spermatogenesis or reproduction. Phylogenetic analysis of *D. melanogaster* S-LAPs and *granny-smith* revealed two distinct S-LAP clusters (Figure 1a): Cluster I includes S-LAPs 3-8 genes (*CG32063*, *CG32064*, *CG18369*, *loopin*, *CG13340*, and

CG4439, respectively) and Cluster II includes S-LAPs 1-2 genes (*CG6372* and *CG32351*, respectively). The average synonymous divergence levels amongst the S-LAP genes ($dS = 1.84$) indicated that the expansion of the gene family was not recent in *Drosophila* evolution. All pairwise values exceeded 1.1 with the exception of *CG6372/CG32351* ($dS = 0.842$), a finding consistent with a more recent gene duplication event resulting in

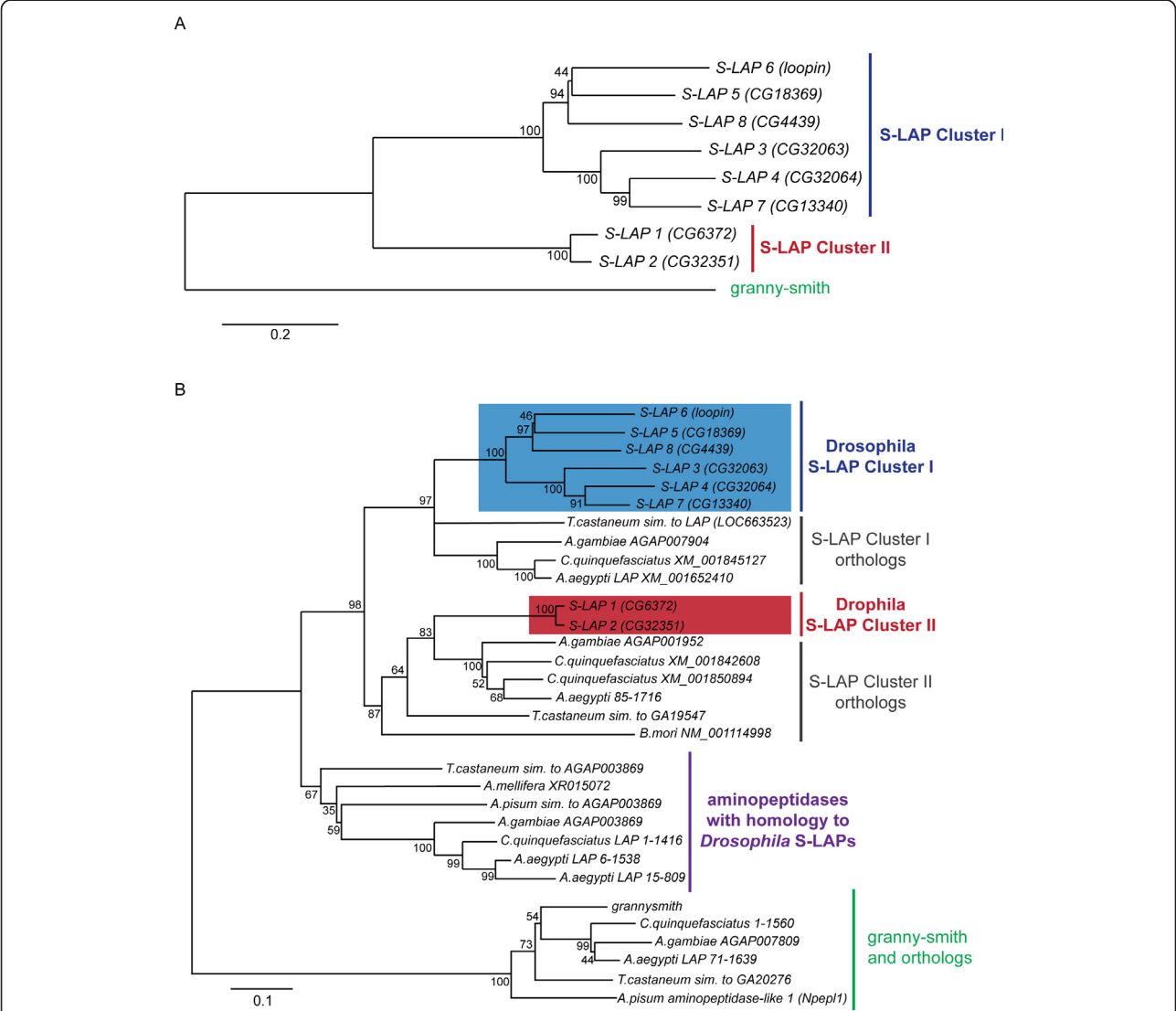


Figure 1 Evolutionary relationship of S-LAPs. (A) Phylogeny of *D. melanogaster* S-LAPs and *granny-smith*. Minimum Evolution distance methods were used to construct the bootstrap consensus protein phylogeny (1000 replicates). Bootstrap values are presented next to relevant nodes and the phylogeny is drawn to scale based on amino acid substitutions per site. This analysis revealed two related clusters of *D. melanogaster* S-LAPS, Cluster I and II. (B) Expanded S-LAP family in *Drosophila*. Comparative genomic analyses identified a total of 17 related aminopeptidases in *A. gambiae*, *C. quinquefasciatus*, *A. aegypti*, *T. castaneum*, *A. pisum* and *A. mellifera*. Minimum evolution distance methods were used to construct the bootstrap consensus protein phylogeny (1000 replicates). Bootstrap values are presented next to relevant nodes and the phylogeny is drawn to scale based on amino acid substitutions per site. S-LAP Cluster I, comprised of 6 genes in *Drosophila*, is related to a single copy aminopeptidase in mosquitos and *T. castaneum*. Similarly S-LAP Cluster II, comprised of two genes in *Drosophila*, is related to a single copy aminopeptidase in other insect taxa with the exception of *C. quinquefasciatus* where two gene copies are present in the genome. In contrast, *granny-smith* has "one-to-one" orthology relationships in other insect taxa.

Cluster II and our phylogenetic analyses (see below). To further estimate the evolutionary timing of S-LAP gene family expansion, orthologs were identified using reciprocal “best-hit” blast searches, for each S-LAP in the 12 available *Drosophila* genomes. This analysis determined that all S-LAPs have unique orthologs in most, if not all the *Drosophila* genomes and that all have distinct orthologs in *D. grimshawi*, the most ancestrally diverged *Drosophila* species for which genomic data is available (a phylogeny containing all identified *Drosophila* orthologs is provided in Additional File 1, Figure S1). Therefore, S-LAP gene family expansion either predates or occurred early in the evolution of the genus *Drosophila*, an observation consistent with the observed levels of synonymous divergence.

Related aminopeptidase genes in other insect taxa

A comprehensive search for related aminopeptidases was conducted in other insect species using reciprocal blast searches against a dataset combining all annotated REFSEQ, non-REFSEQ and *ab initio* gene sequences in *Anopheles gambiae*, *Culex quinquefasciatus*, *Aedes aegypti*, *Tribolium castaneum*, *Nasonia vitripennis*, *Acyrtosiphon pisum*, *Bombyx mori* and *Apis mellifera*. This resulted in the identification of 17 aminopeptidases related to the S-LAP family and 6 orthologs of *granny-smith*. The number of identified S-LAP related aminopeptidase genes varied between taxa, ranging from none in *N. vitripennis* to 4 in *A. aegypti* and *A. gambiae*. Copy number determination of related aminopeptidase genes in each species was also confirmed through the analysis of genome assembly sequences in each species (data not shown).

S-LAPs: an expanded gene family in *Drosophila*

Phylogenetic analysis of the S-LAP gene family, including related aminopeptidases from other taxa, revealed that S-LAP Cluster I and II are related to single-copy genes in most insect taxa (Figure 1b). The 6 members of Cluster I form a monophyletic group with a single related gene in all 3 mosquito species and the beetle,

T. castaneum. This is consistent with gene family expansion during early *Drosophila* evolution. Furthermore, no Cluster I related S-LAP genes were detected in *N. vitripennis*, *A. pisum*, *B. mori* and *A. mellifera*. The two more closely related *Drosophila* members of Cluster II also form a monophyletic group with single-copy genes in other insect taxa with the exception of *C. quinquefasciatus* where two related aminopeptidases were found. These observations are in contrast to *granny-smith* that has “one-to-one” orthology relationships in all taxa where an ortholog was identified. It is also noteworthy that our analysis identified a set of seven aminopeptidases in non-*Drosophila* taxa (Figure 1b) distinct from either *Drosophila* S-LAP Cluster I or II. This includes a single copy aminopeptidase gene across 5 insect taxa and 2 putative, additional paralogs in *A. aegypti*.

Mechanisms of new S-LAP gene creation

The age of S-LAP clusters and the associated sequence divergence at these loci means that hallmarks associated with gene duplication mechanisms are often obscured, making gene duplication mechanisms difficult to infer. However, two pairs of S-LAPS (S-LAP 1 and 2; S-LAP 3 and 4) are immediately adjacent in the genome and share common intron/exon structure suggesting they were created through tandem duplication. S-LAP 7 has previously been characterized as a retrotransposed gene [11,12]. Consistent with a retroposition event, S-LAP 7 is intronless, unlike its likely progenitor S-LAP 4, and these genes reside in different genomic regions (2R: 50C5 and 3L: 67E4, respectively).

S-LAPs are expressed specifically in the testis

Expression of each S-LAP in *D. melanogaster* was characterized by RT-PCR using gene specific primers. A high level of testis-specific expression for all eight S-LAPs was observed with no detectable expression in gonadectomized males (Figure 2), a finding consistent with previous microarray studies [35]. The testis-specific expression of the S-LAP family is in contrast to *granny-smith*, the closest related leucyl aminopeptidase in the



Figure 2 S-LAPs are specifically expressed in testis. RT-PCR analysis reveals testis specific expression of the *D. melanogaster* S-LAPs. Testis expression is compared to expression in the remainder of the fly using gonadectomized males.

genome, which is expressed broadly across a range of tissues, including (but not limited to) testis, ovary, tubule and salivary gland [35].

Confirmation and quantification of S-LAPs in *Drosophila* Sperm by 2D-Gel

As part of an exhaustive analysis of sperm proteins from 2D-gels of varying isoelectric gradients, all eight S-LAPs were identified by both MALDI-TOF and tandem MS/MS analyses (Figure 3, Table 1). High confidence of S-LAP protein identification was supported by, (i) multiple peptide assignments for each protein (with an

average of 5.8 peptides per S-LAP protein) as assigned by Scaffold (Proteome Sciences); (ii) Mowse scores from Maldi-tof measurements that exceeded statistical scoring requirements [36] and (iii) replicate spot assignments from multiple 2D-gels under varying separation conditions (Figure 3). Tubulins were also identified on 2D-gels at the expected molecular weight and isoelectric point positions (Figure 3).

Visual inspection of 2D-gels clearly indicated that both the S-LAPs and tubulins represent a substantial proportion of the total protein in sperm on 2D-gels (brackets, Figure 3). We determined the relative abundance of

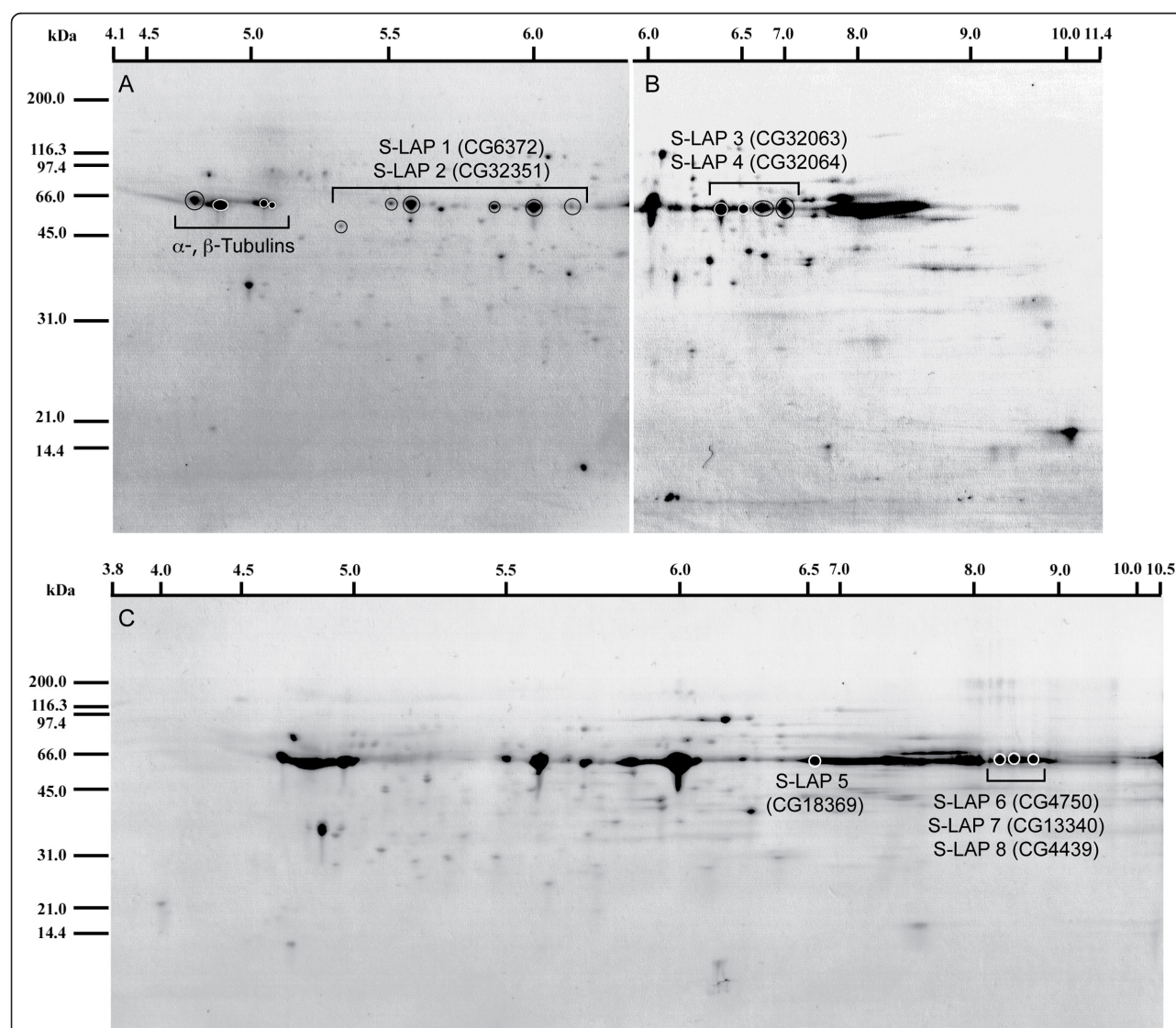


Figure 3 Annotation of S-LAPs on 2D gels of purified sperm. Gels were run using duplicate samples of solubilized sperm protein (100 ug/gel). Proteins were first separated on isoelectric focusing gels with pH ranges of (A) 4-7, (B) 6-11 and (C) 3-10 followed by second dimension SDS-PAGE using 12% Bis-Tris reducing gels. Spots were visualized using Coomassie Blue R250 and spots excised and submitted for MALDI-TOF and tandem MS/MS analyses. The α- and β-tubulins are indicated in (A) at the expected pI and molecular weight positions. Regions of the gel identified as S-LAPs are indicated by brackets with individual spots circled.

Table 1 S-LAPs identified in *D. melanogaster* sperm by MS

Gene Name	FlyBase ID	Chromosome (Cytological Location)	Peptide Fragments Identified by Whole Sperm Tandem MS (# of times identified)*	Confirmed Mascot ID**
S-LAP 1 (CG6372)	FBgn0035915	3L (66D8-66D8)	Y.YSWAVVGLGKE R.FPLWNFYSKA (2) R.LLFAVEPERI R.IFVPAIDLATK.D	Yes
S-LAP 2 (CG32351)	FBgn0052351	3L (66D8-66D8)	R.GDEFEYLRE (2) R.ISVPSIDLYATK.D (4) R.VPYSAVAVIGLGKE R.FPLWNYYSKA	Yes
S-LAP 3 (CG32063)	FBgn0045770	3L (67E3-67E4)	K.VFNNVDSEFR.S K.GIMLGLYEKE R.IAAGIGAR.S K.GIMLGLYEKE K.LELHGSPPVESWTR.G (4)	Yes
S-LAP 4 (CG32064)	FBgn0052064	3L (67E4-67E4)	R.LWVDATLGSGVK.K K.QITDEMGYDLSNDGR.G (2) K.QFQSAGALTGDR.V K.IFNNIDPEFR.S K.PGDVVTLMNHK.S K.GLVGVYQKE K.LELYESPDYEGWTR.G (2) K.GITFNSGAMNLR.K	Yes
S-LAP 5 (CG18369)	FBgn0033860	2R (50B4-50B5)	R.HNPLPYLLKDR.M R.VAAGVGARA R.ETGIKGLGK.G K.QLVTPNLTFDISNR.G	Yes
S-LAP 6 (loopin)	FBgn0259795	2R (53C9-53C10)	K.QGAAYNENEELDEGMENVR.V (5) N.AVTLDDALGGKL R.GTGMTTINPR.P K.TTANAVTLDDALGGKL (3) K.GWVGVTYTKD	Yes
S-LAP 7 (CG13340)	FBgn0033868	2R (50C5-50C5)	R.LWDIATLNTGVKK (2) R.LPLWQYYR.R K.YGLVPYLTKK K.GVWGLYQKE	Yes
S-LAP 8 (CG4439)	FBgn0034132	2R (53C6-53C6)	K.LELYGSTDQDAWTR.G (5) K.HGVPPYLLK.D (3) R.LYQNIDKE K.ESGLNGELGVGR.L K.EGAGFNAEEVIDEGMENVR.V (2)	Yes

*MS analysis from [18].

** 2D-gel protein spot locations presented in Figure 3.

each by comparing spot signal intensity for each class. Five 2D-gels (pI range 4-7, Figure 2) were processed, scanned and protein spot intensities calculated. The results indicate an approximate 2-fold abundance of S-LAPs (Table 2) over tubulins. Thus, we conclude that the tubulins and the S-LAP family represent the major mass of protein in sperm.

S-LAP enzymatic activity in purified sperm

S-LAP activity in sperm extracts was directly measured using L-Leu-AMC, a sensitive and specific fluorogenic substrate for leucine aminopeptidase activity [37]. Negligible activity was observed in intact sperm suggesting that S-LAPs were not exposed on the sperm surface (L-Leu-AMC is not permeable to intact cells). Sonication of

sperm in buffer alone resulted in measurable S-LAP activity but with low protein yield (Table 3). Addition of DTT resulted in both higher levels of soluble protein and measurable S-LAP activity but with reduced specific activity (Table 3). However, while this establishes presence of leucylaminopeptidase activity, the relative contribution of each S-LAP to the observed signal remains unclear.

Divergence at cation ligand and catalytic residues

Biochemical and x-ray crystallography studies have defined 7 primary active site residues in M17 leucyl aminopeptidases involved in catalysis and binding of divalent cations, including a critical Zn^{2+} ion binding region [38,39]. These sites are highly conserved in leucyl aminopeptidases across animal taxa strongly suggesting their

Table 2 Relative levels of tubulins and S-LAPs from 2D gels

Spot Intensity (arbitrary units $\times 10^6$)						
2D-gel	Tubulin	S-LAPs	Gel Total	Tubulin/total (%)	S-LAP/total (%)	LAP/Tubulin
Gel 1	2.42	5.00	50.40	4.80	9.93	2.07
Gel 2	1.42	3.23	39.50	3.58	8.19	2.28
Gel 3	2.03	3.68	43.26	4.69	8.51	1.81
Gel 4	2.01	3.88	52.10	3.87	7.44	1.92
Gel 5	1.90	5.36	66.02	2.89	8.12	2.81
Average				3.97	8.44	2.18
Median				3.87	8.19	2.07
S. D. (+/-)				0.80	0.92	0.34

critical role in the coordination of the metal ions, water and a bicarbonate ion which coordinates a base attack on the peptide bond followed by breakdown of a tetrahedral diol intermediate with the terminal amino acid as the leaving group [40]. Intriguingly, the S-LAP active site has diverged during *Drosophila* evolution replaced in many cases by amino acids incapable of Zn^{2+} binding or bicarbonate coordination (Figure 4). In contrast, *granny-smith* contains the canonical active site residues. Furthermore, differences with likely functional relevance exist between the active site residues in S-LAP clusters I and II. The two members of Cluster II retain a high level of amino acid identity with the M17 aminopeptidase consensus, including i) 3 out of 4 conserved amino acids amongst the tight Zn^{2+} binding residues (commonly referred to as “site 2”), ii) a conservative amino acid substitution (Lys327Gln) to a known M17 metal ligand amino acid at the fourth site and iii) a conservative amino acid substitution (Lys409Asp) to a known M17 metal ligand amino acid which contributes to divalent cation binding at the “loose” binding site (commonly referred to as “site 1”). This is in contrast to S-LAPs in Cluster I that possess i) limited or no similarity to the M17 aminopeptidase consensus and ii) an abundance of substitutions in residues that do not bind metal ions or are not characterized as M17 LAP metal ion ligands. Furthermore, 3 of the 6 S-LAPs in Cluster I have undergone a substitution at Lys339 which is essential for enzymatic activity [41]. The 7 characterized functional

residues of the active site are displayed on the S-LAP protein alignment in Additional File 2, Figure S2.

Purifying selection on the S-LAP gene family

Given the rapid expansion of the S-LAP gene family and the potential loss of catalytic activity amongst S-LAPs, we next sought to assess the role of positive selection in the functional diversification of the S-LAP gene family. To this end, we conducted codon based maximum likelihood analyses of selection for S-LAP Cluster I and II using all *Drosophila* S-LAP orthologs. This analysis resulted in strong evidence for purifying selection throughout the S-LAP gene. Analysis of all Cluster I sequences (including 68 gene sequences) resulted in a mean dN/dS estimate of 0.14, significant evidence of negative selection ($p < 0.05$) at 455 sites and no codons showing evidence of being under positive selection. Similarly, analysis of Cluster II (including 18 genes) resulted in a mean dN/dS estimate of 0.12, significant evidence of negative selection at 425 sites and no codons showing evidence of being under positive selection.

Discussion

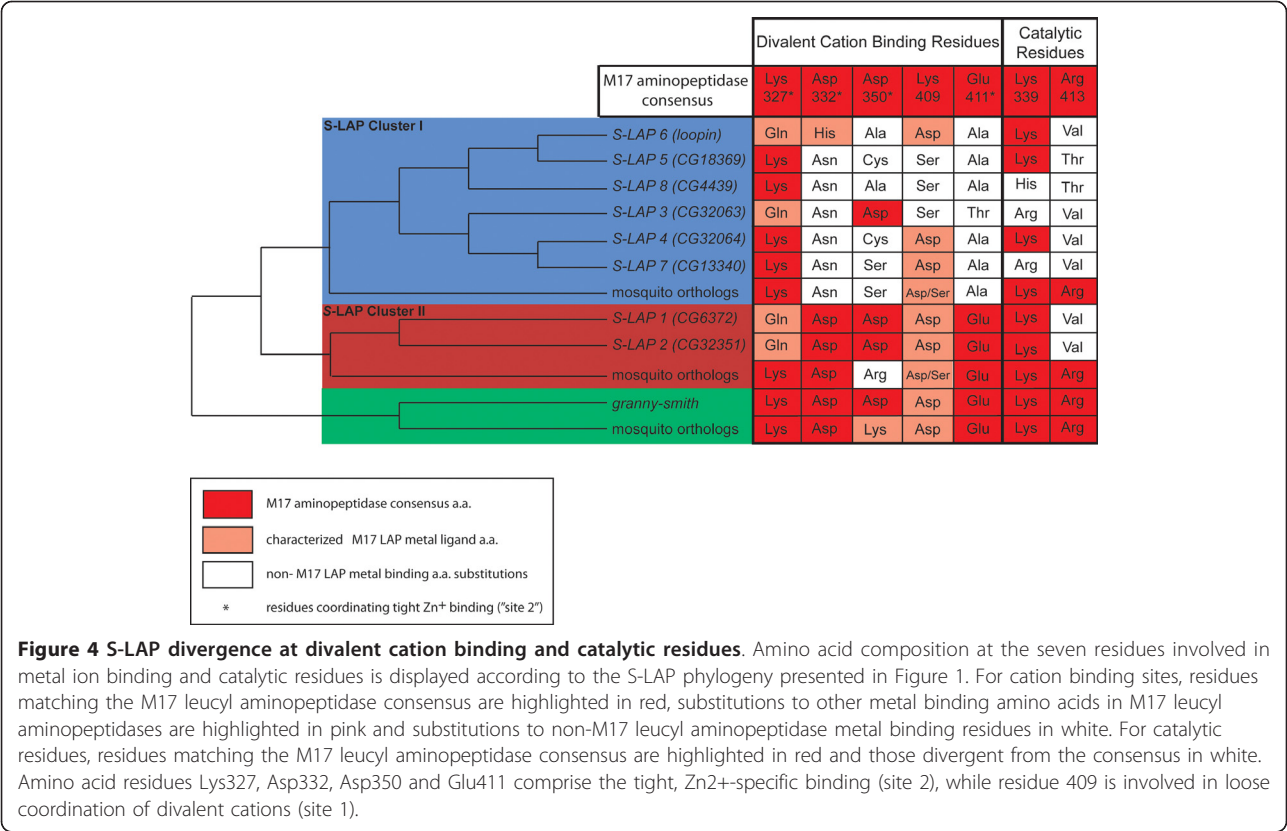
Advances in comparative genomics have revealed that the creation of novel genes is a widespread process [11,42-44] and that the composition of gene families associated with reproduction can vary dramatically between species [10]. Often, the expansion of gene families through gene creation results in functional diversification of gene family members, a process sometimes associated with rapid evolutionary changes at the DNA sequence and spatio-temporal expression levels. Furthermore, it has also been well documented that new retrogenes often acquire testis expression, however a limited amount is known about the specific functional ramifications of these phenomenon. To map specific functional consequences of gene family evolution, we have utilized comparative genomics to characterize the dynamic evolutionary history of the S-LAP gene family and documented S-LAP abundance and activity in spermatozoa.

Table 3 Leucine aminopeptidase activity in purified *Drosophila* sperm

Sperm Treatment	Protein (ug)	% total*	Activity (mU)	S.A.‡ (U/mg)
Buffer	<0.25	0.6	0	0
Buffer + sonication	1.1	2.2	1.78	1.62
Buffer+DTT +sonication	5.4	10.8	4.39	0.81

*The percent values are calculated from maximal amount of extractable sperm protein.

‡ Specific activity.



Gene family expansion and evolution

The recent availability of additional insect genome sequences has allowed the analysis of gene family evolution as it potentially relates to specific biological characteristics [45]. This approach has been of great utility in our studies of *Drosophila* sperm where gene creation has proven to be a widespread process and potentially a dynamic force in the molecular evolution of sperm [12]. Comparative genomic analyses of the sperm-specific S-LAPs in this study further extends this observation and demonstrates that the S-LAP gene families are highly heterogeneous in composition across the insect genomes that have been sequenced to date. This includes, but is not limited to, a dramatic expansion during the evolution of *Drosophila* (especially within S-LAP Cluster I) and variable copy numbers across three mosquito species. While some revision of the S-LAP gene family evolution may result from the availability of additional insect genomes, these results clearly reiterate the fundamental importance of gene creation during the evolution of reproductive traits in *Drosophila*.

Leucyl aminopeptidases in *Drosophila* sperm

Aminopeptidases, a large family of related enzymes found throughout animal and plant kingdoms, have been described in many reproductive tissues including

sperm and seminal fluids [46-48]. Our results extend these observations to include a newly recognized family, S-LAPs 1-8. Other than a role for a GPI-anchored aminopeptidase in the initiation of the acrosome activation in mussels (*Mytilus edulis*) [49], little functional data exists for leucylaminopeptidases either in sperm or indeed other aspects of reproductive biology. Our results provide strong circumstantial data suggesting that S-LAPs may play an important role in sperm function: (i) S-LAPs are expressed exclusively during spermatogenesis; (ii) they collectively represent the most abundant protein class in *Drosophila* sperm by a 2:1 mass ratio compared to tubulin (Figure 3; Table 2); (iii) bioinformatic analysis demonstrated S-LAPs as the most significantly over-represented protein domain within the proteome [33]; and (iv) analysis of stage specific expression during spermatogenesis (i.e. mitotic, meiotic and post-meiotic) revealed that all 8 S-LAPs were upregulated specifically during the meiotic and post-meiotic stages of late spermatogenesis (data not shown) [50], the developmental time window of spermiogenesis and sperm production. Further, we observed significant levels of leucyl aminopeptidase activity in both the soluble and insoluble fractions of purified sperm. S-LAP activity was recovered from the insoluble pellet in the presence of DTT. Furthermore, the two fractions

differed in specific activities perhaps reflecting a partition of different S-LAPs that may be directly related to functional diversification between them (see below).

In contrast to the rapid evolution often observed in genes associated with male reproduction, we observe a strong signal of purifying selection across all S-LAPs suggesting a constrained functional role in sperm. This conservation is intriguing given that each cluster displays distinct patterns of substitutions at the active site consistent with functional diversification. Cluster I has undergone radical amino acid replacements in many of the canonical active site residues and all model studies of this class of enzyme suggest that these substitutions would be likely to render the catalytic site inactive. This observation, in conjunction with the evolutionary conservation (presumably due to functional constraint) of Cluster I, suggests the possible evolutionary acquisition of important functions independent of enzymatic activity. Comparison of Cluster I active site residues in *Drosophila* and mosquito reveals that most active site substitutions occurred prior to the evolutionary expansion of Cluster I, indicating that the Cluster I progenitor gene had potentially lost enzymatic activity prior to the expansion of this cluster (Figure 4). Future studies are needed to determine if the mosquito orthologs of Cluster I also function specifically in sperm.

In contrast, Cluster II S-LAPs more closely match the canonical M17 aminopeptidase active site and are likely to account for the enzymatic activity in purified sperm quantified in this study. The amino acid substitution (Arg413Val) in the catalytic site occurs at a residue not strictly required for enzymatic activity but instead is expected to decrease activity by 10 to 100-fold based on previous biochemical studies [51,52]. Inspection of the mosquito Cluster II active sites indicates that this substitution occurred more recently during *Drosophila* evolution. The other catalytic site, residue Lys339, is strictly conserved and has been shown essential for catalytic activity [41]. It is possible that the CG6372/CG32351 duplication event may have partially restored levels of enzymatic activity lost as a result of the Arg413Val substitution. Quantification of specific activity for each S-LAP is the next necessary step in assessing the functional diversification of the gene family.

Conclusions

The presence of eight distinct S-LAPs and their extraordinarily high protein concentration in sperm strongly suggests that S-LAPs play a predominant (as yet unknown) role in *Drosophila* sperm biology. However, our results suggest that this role (or roles) cannot be linked to enzymatic activity in Cluster I S-LAPs where the canonical active site is clearly absent. An outstanding question therefore remains: what is the acquired

functionality of Cluster I S-LAPs? One testable possibility is that Cluster I S-LAPs lost enzymatic activity but were retained, and subsequently diversified, following the acquisition of a structural function in sperm. While such a scenario for sperm protein evolution is unprecedented, it has been well documented in other enzyme families such as the crystallin family of lens proteins that evolved from ubiquitously expressed metabolic enzymes (reviewed in [53]). Although the precise evolutionary processes leading to the neofunctionalization of the crystallins are unknown, it did include the loss of enzymatic activity due to substitutions at critical active site residues, the expression of proteins at high concentrations and the subsequent establishment of their current structural and refractive roles in the eye lens [53]. It is noteworthy that the S-LAPs share extensive sequence homology (>40% at the amino acid level) to the bovine lens leucine aminopeptidase (data not shown). It will be of great interest to determine if, and to what extent, these two independent evolutionary processes parallel one another. Regardless, the S-LAP gene family represents a unique example wherein loss of function (and the presumed gain of a novel function) predates the dramatic expansion of a gene family. As such, neofunctionalization may itself lead to a situation where gene family expansion is selectively favorable, an observation that complements the general model where gene duplication predates and facilitates the evolution of novel functionality. Finally, in recent years characterized functions of LAPs have expanded to include a range of other cellular processes (e.g., transcriptional regulation, site-specific recombination and meiosis; reviewed in [54]). It will be of considerable interest to investigate whether S-LAPs have similarly acquired diversified roles in sperm function and physiology.

Methods

Comparative genomic analysis of the S-LAP family

BlastN was utilized to identify the closest paralog of the S-LAP gene family (Additional File 3), *granny-smith*, amongst annotated *D. melanogaster* genes and confirmed by estimates of synonymous divergence (dS). S-LAP orthology amongst *Drosophila* species was determined using reciprocal "best-hit" blastP against all annotated *Drosophila* proteins and was found to be consistent with current orthology annotation tables (<http://flybase.org/>). Identification of aminopeptidase genes related to the S-LAPs in other insects was conducted using a reciprocal blast approach where the eight S-LAPs were first compared to a combined database of RefSeq, non-RefSeq and ab initio coding sequences from *A. gambiae*, *C. quinquefasciatus*, *A. aegypti*, *T. castaneum*, *N. vitripennis*, *A. pisum* and *A. mellifera* using Tblastx. The second, reciprocal search compared all matches from the first search to all annotated *D. melanogaster* genes. Those that best

matched members of the S-LAP family (in any order) were kept for further analysis. The same procedure was used to identify orthologs of S-LAP paralog, *granny-smith*. Concordant results, including a consistent estimate of S-LAP related gene copy number in each insect genome, were obtained using TblastN against the whole genome assembly for each species.

Phylogenetic and evolutionary analysis

Alignments were conducted using CLUSTALW and phylogenetic analyses were conducted using the Minimum Evolution Method as implemented by MEGA4 [55]. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history. Distances were calculated as the number of amino acid substitutions per site using the Poisson correction method. Consistent phylogenies were obtained using Bayesian phylogenetic methods [56]. Synonymous (dS) and nonsynonymous (dN) levels of divergence were estimated from "in-frame" CLUSTALW nucleotide alignments using the method of Nei and Gojobori [57]. Codon based analyses of selection were conducted using maximum-likelihood methods as implemented by HyPhy [58].

RT-PCR analysis

Male carcasses (minus gonads) and testes were dissected, washed (3x) in phosphate buffered saline and flash frozen in liquid nitrogen. Total RNA was then isolated using manufacturer's protocols (Ambion195 RNA-queous Kit), and RNA was suspended in TE and quantified. Complementary DNA (cDNA) was carried out by standard protocols (Promega Improm-II Reverse Transcription System) using 75-ng input RNA (final concentration 3.75 ng/ul) from whole males (minus gonads) and testis. Gene-specific polymerase chain reaction (PCR) primers were used to amplify each S-LAP transcript from equal quantities of cDNA from the whole fly (minus gonads) and testis. Primer sequences are provided in Additional file 4. Equal volumes of each PCR reaction were visualized on 1% agarose gels.

Sperm sample preparation and protein quantification

D. melanogaster sperm purified from Tempe-T strain males was solubilized over night at 25°C in CHAPS buffer containing 7 M Urea, 2 M Thiourea, 4% CHAPS, 33 mM DTT and 2.5 mM TCEP). The solubilized protein was separated from insoluble aggregates by ultracentrifugation, using a Beckmann-Coulter TL-100 and a TLA 100.2 fixed angle rotor for 30 min, 88,760 × g at 17°C. The supernatant was transferred in a new tube and stored at -80°C. Protein quantitation was measured using the RediPlate Protein Quantitation Kit (Molecular Probes) and calibration curves as per manufacturers instructions and fluorescence signals determined using a

Fuji Laser Scanner FLA5000. Protein values were calculated using the Aida 2D-Densitometry software (Raytest, Straubenhardt, Germany).

Protein precipitation for 2D gel electrophoresis

Sperm extracts were precipitated with 100% ice cold TCA (Sigma) at a final concentration of 20%. Following a 10 min incubation on ice, precipitated protein was collected by centrifugation for 5 min, 16 000 × g at 4°C. The supernatant was discarded and the pellet was washed three times with ice-cold acetone. After removing the acetone by a brief incubation at 37°C, the pellet was suspended in CHAPS buffer.

2D-Gel electrophoresis

The Multiphor II system from GE Healthcare was used for isoelectric focusing and SDS polyacrylamide gel electrophoresis following the manufacturers guidelines. Immobiline DryStrip gels 11 cm (pH 4-7, 6-11) and 24 cm (3-10NL) and precast 12.5% homogenous SDS gels were purchased from GE Healthcare. Immobiline Dry strip gels were rehydrated overnight in rehydration buffer (200 ml for 11 cm and 450 ml for 24 cm) containing 7 M Urea, 2 M Thiourea, 4% CHAPS, 18 mM DTT, 0.5% IPG buffer. pH4-7 and 3-10 NL Immobiline Dry strip gels were hydrated directly with the protein sample in rehydration buffer. Isoelectric focusing was performed at 20°C using the following 3-phase run parameters (upper limits for current and power were set to 2 mA and 2 W, respectively): phase 1, 300 V for 1 volt-hour (Vh); phase 2, a linear voltage increase to 3500 V over 2900 Vh; phase 3, constant 3500 V for 9.1 kVh (pH4-7) or 6.1 kVh (pH6-11). The isoelectric focusing protocol for the 24 cm 3-10NL; phase 1, 500 V for 1 Vh; phase 2, linear voltage increase to 3500 V over 3000 Vh; phase 3, constant 3500 V for 57 kVh. The strips were either frozen at -80°C or used immediately for second dimension gel electrophoresis. Before SDS gel electrophoresis strips were equilibrated for 15 min in 50 mM Tris pH 8.8, 6 M Urea, 30% Glycerol, 2% SDS, 0.002% bromphenolblue and 65 mM DTT to reduce disulfide bonds. A second equilibration step using the same reagents as above, but the DTT was replaced with 135 mM iodoacetamide. Iodoacetamide alkylates thiol groups prevents oxidation during electrophoresis. The Immobiline DryStrip gel was then transferred gel side down onto the SDS polyacrylamide gel. Gels were washed overnight in a 40% ethanol/10% acetic acid fixation solution. Proteins were visualized using Phastgel Blue Coomassie R350 (GE Healthcare) following manufacturer protocol.

Mass spectrometry identification of S-LAPs

Eight S-LAPs were previously identified in the *Drosophila* sperm proteome by high throughput tandem

mass spectrometry of whole sperm digests [18] and confirmed in a recent shotgun proteomic analysis [33]. The 2D-gel migration patterns of all eight S-LAPs were determined in an exhaustive study that identified the molecular identity of >80 major protein spots (U. Gerike and TLK, unpublished). Spots were excised from the gel and transferred into Microtiter plates, destained and digested using the Ettan Digester (GE Healthcare). The proteins were digested with sequencing grade trypsin (Promega) at 20 ng trypsin/ug sperm protein according to the manufacturers protocol. The resulting peptide mixtures were mixed with a saturated solution of α -cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% trifluoroacetic acid, spotted and dried prior to analysis. Peptide masses were determined using a MALDI-TOF mass spectrometer (Micromass) running in linear mode. These results were further validated using tandem mass spectrometry (Orbitrap, Thermo Inc.). Peptides, prepared as described above, were subjected to tandem MS/MS and spectra analyzed using Scaffold (v2.02, Proteome Software, Portland, Oregon, USA). Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm [59] and at least two tryptic peptides were detected. The average number of peptides identified per S-LAP protein was 5.8.

Analyzing the MALDI TOF results

The Mowse scoring algorithm in a probabilistic framework of the Mascot Peptide Mass Fingerprint software (Matrix Science) was used as search tool. The search criteria were as follows: database - MSDB; organism - *Drosophila*; protease - Trypsin; fixed modification, Carbamidomethyl; variable modification - Oxidation (M); mass accuracy - 100 ppm and 0 missed cleavages.

Relative protein quantification of gel spots

Three coomassie stained 2D-gels from *D. melanogaster* sperm in the pH 4-7 were scanned using FLA-5000 phosphorimager in digital mode and using a 472 nm excitation filter. Relative spot intensities were measured using version 3.23 of the AIDA 2D densitometry software (Raytest, Straubenhardt, Germany). The LAU (light absorbing units) were determined for each spot with background correction.

Protein extraction and S-LAP activity assays

Sperm from 25 males of flies were incubated for 1 h at 25°C in 500 ml buffer A (50 mM HEPES pH7.4, 1 mM MnCl₂, 67.5 mM DTT). Sperm was then broken up by sonication (~6s on ice) and the sample centrifuged 30 min at 16 000 g at 4°C. The supernatant was transferred

to a new tube. The incubation mixture contained (in a total of 3 ml): 50 mM HEPES pH7.4, 1 mM MnCl₂, 11.5 mM DTT and 100 mM L-Leu-AMC [Sigma]. S-LAP activity was determined by measuring the increase in fluorescence emission at 460 nm (340 nm excitation) using a Luminescence Spectrometer LS50B (Perkin Elmer). The reaction was started by the addition of the sperm extract. To calculate the units of activity a calibration curve was performed with 7-amino-4-methylcoumarin [Sigma] in the range 0-7 nmol.

Additional material

Additional file 1: Complete S-LAP phylogeny. This file includes a complete S-LAP phylogeny with all *Drosophila* orthologs.

Additional file 2: S-LAP protein alignment. This file includes the S-LAP protein alignment denoting the seven characterized functional residues of the active site.

Additional file 3: S-LAP primers. This file includes a table listing the primers used in S-LAP RT-PCR experiments.

Additional file 4: Name and symbol key. This file includes a table listing the names and symbols of the genes encoding aminopeptidases referred to in this study.

Acknowledgements

The authors wish to thank U. Gerike for 2D gel and Maldi-tof analyses and Jean van den Elsen for expert technical advice. This research was supported by awards from the Royal Society (Wolfson Merit Award), and the Biodesign Institute, Arizona State University (TLK). Additional support was provided by a Ruth L. Kirschstein National Research Service Award (National Institutes of Health), and an Academic Research Fellowship from the Research Council of the United Kingdom (SD).

Author details

¹Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK. ²Centers for Evolutionary Medicine and Informatics and Infectious Diseases and Vaccinology, Biodesign Institute, Arizona State University, Tempe, AZ 85287-5001, USA.

Authors' contributions

TLK and SD conceived the experiments and wrote the paper and ECW performed the PCR experiments. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 15 November 2010 Accepted: 5 April 2011

Published: 5 April 2011

References

- Ohno S: *Evolution by gene duplication*. Springer, Berlin; 1970, 1970.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: *Preservation of duplicate genes by complementary, degenerative mutations*. *Genetics* 1999, **151**:1531-1545.
- Lynch M, Force A: *The probability of duplicate gene preservation by subfunctionalization*. *Genetics* 2000, **154**:459-473.
- Lynch M, Conery JS: *The evolutionary fate and consequences of duplicate genes*. *Science* 2000, **290**:1151-1155.
- Maston GA, Ruvolo M: *Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection*. *Mol Biol Evol* 2002, **19**:320-335.

6. Brown CJ, Todd KM, Rosenzweig RF: **Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment.** *Mol Biol Evol* 1998, **15**:931-942.
7. Dai HZ, Chen Y, Chen SD, Mao QY, Kennedy D, Landback P, Eyre-Walker A, Du W, Long MY: **The evolution of courtship behaviors through the origination of a new gene in Drosophila.** *Proc Natl Acad Sci USA* 2008, **105**:7478-7483.
8. Loppin B, Lepetit D, Dorus S, Couble P, Karr TL: **Origin and neofunctionalization of a Drosophila paternal effect gene essential for zygote viability.** *Curr Biol* 2005, **15**:87-93.
9. Kalamegham R, Sturgill D, Siegfried E, Oliver B: **Drosophila mojoless, a retroposed GSK-3, has functionally diverged to acquire an essential role in male fertility.** *Mol Biol Evol* 2007, **24**:732-742.
10. Hahn MW, Han MV, Han SG: **Gene family evolution across 12 Drosophila genomes.** *PLoS Genet* 2007, **3**:e197.
11. Betran E, Thornton K, Long M: **Retroposed new genes out of the X in Drosophila.** *Genome Res* 2002, **12**:1854-1859.
12. Dorus S, Freeman ZN, Parker ER, Heath BD, Karr TL: **Recent origins of sperm genes in Drosophila.** *Mol Biol Evol* 2008, **25**:2157-2166.
13. Emerson JJ, Kaessmann H, Betran E, Long M: **Extensive gene traffic on the mammalian X chromosome.** *Science* 2004, **303**:537-540.
14. Dorus S, Gilbert SL, Forster ML, Barndt RJ, Lahn BT: **The CDY-related gene family: coordinated evolution in copy number, expression profile and protein sequence.** *Hum Mol Genet* 2003, **12**:1643-1650.
15. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H: **Emergence of young human genes after a burst of retroposition in primates.** *PLoS Biol* 2005, **3**:e357.
16. Vinckenbosch N, Dupanloup I, Kaessmann H: **Evolutionary fate of retroposed gene copies in the human genome.** *Proc Natl Acad Sci USA* 2006, **103**:3220-3225.
17. Nurminsky DI, Nurminskaya MV, Benevolenskaya EV, Shevel'yov YY, Hartl DL, Gvozdev VA: **Cytoplasmic dynein intermediate-chain isoforms with different targeting properties created by tissue-specific alternative splicing.** *Mol Cell Biol* 1998, **18**:6816-6825.
18. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL: **Genomic and functional evolution of the Drosophila melanogaster sperm proteome.** *Nat Genet* 2006, **38**:1440-1445.
19. Wigby S, Chapman T: **Sex peptide causes mating costs in female Drosophila melanogaster.** *Curr Biol* 2005, **15**:316-321.
20. Ravi Ram K, Sirot LK, Wolfner MF: **Predicted seminal astacin-like protease is required for processing of reproductive proteins in Drosophila melanogaster.** *Proc Natl Acad Sci USA* 2006, **103**:18674-18679.
21. Wong A, Albright SN, Giebel JD, Ram KR, Ji S, Fiumera AC, Wolfner MF: **A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in Drosophila melanogaster.** *Genetics* 2008, **180**:921-931.
22. Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW: **Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (Mus domesticus).** *Mol Biol Evol* 2009, **26**:1733-1743.
23. Wolfner MF: **"S.P.E.R.M." (seminal proteins (are) essential reproductive modulators): the view from Drosophila.** *Soc Reprod Fertil* 2007, **65**:183-199.
24. Kelleher ES, Pennington JE: **Protease gene duplication and proteolytic activity in Drosophila female reproductive tracts.** *Mol Biol Evol* 2009, **26**:2125-2134.
25. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT: **Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity.** *Nat Genet* 2004, **36**:1326-1329.
26. Swanson WJ, Clark AG, Waldrup-Dail HM, Wolfner MF, Aquadro CF: **Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila.** *Proc Natl Acad Sci USA* 2001, **98**:7375-7379.
27. Parker GA: **Selection on non-random fusion of gametes during the evolution of anisogamy.** *J Theor Biol* 1978, **73**:1-28.
28. Rice WR: **Sexually antagonistic male adaptation triggered by experimental arrest of female evolution.** *Nature* 1996, **381**:232-234.
29. Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL: **Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes.** *Mol Biol Evol* 2010, **27**:1235-1246.
30. Baker MA, Hetherington L, Reeves GM, Aitken RJ: **The mouse sperm proteome characterized via iPG strip prefractionation and LC-MS/MS identification.** *Proteomics* 2008, **8**:1720-1730.
31. Stein KK, Go JC, Lane WS, Primakoff P, Myles DG: **Proteomic analysis of sperm regions that mediate sperm-egg interactions.** *Proteomics* 2006, **6**:3533-3543.
32. Rawlings ND, Barrett AJ, Bateman A: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2010, **38**:D227-33.
33. Wasbrough R, Dorus S, Hester S, Howard-Murkin J, Lilley KS, Wilkin E, Polpitiya A, Petritis K, Karr TL: **The Drosophila melanogaster sperm proteome-II (DmSP-II).** *J Proteomics* 2010, **73**:2171-2185.
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
35. Chintapalli VR, Wang J, Dow JA: **Using FlyAtlas to identify better Drosophila melanogaster models of human disease.** *Nat Genet* 2007, **39**:715-720.
36. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20**:3551-3567.
37. Saifuku K, Sekine T, Namihisa T, Takahashi T, Kanaoka Y: **A novel fluorometric ultramicro determination of serum leucine aminopeptidase using a coumarine derivative.** *Clin Chim Acta* 1978, **84**:85-91.
38. Kim H, Burley SK, Lipscomb WN: **Re-refinement of the X-ray crystal structure of bovine lens leucine aminopeptidase complexed with bestatin.** *J Mol Biol* 1993, **230**:722-724.
39. Maric S, Donnelly SM, Robinson MW, Skinner-Adams T, Trenholme KR, Maric S, Donnelly SM, Robinson MW, Skinner-Adams T, Trenholme KR, Gardiner DL, Dalton JP, Stack CM, Lowther J: **The M17 leucine aminopeptidase of the malaria parasite Plasmodium falciparum: importance of active site metal ions in the binding of substrates and inhibitors.** *Biochemistry* 2009, **48**:5435-5439.
40. Strater N, Sun L, Kantrowitz ER, Lipscomb WN: **A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase.** *Proc Natl Acad Sci USA* 1999, **96**:11151-11155.
41. Strater N, Sherratt DJ, Colloms SD: **X-ray structure of aminopeptidase A from Escherichia coli and a model for the nucleoprotein complex in Xer site-specific recombination.** *EMBO J* 1999, **18**:4513-4522.
42. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al: **Evolution of genes and genomes on the Drosophila phylogeny.** *Nature* 2007, **450**:203-218.
43. Wang W, Zhang J, Alvarez C, Llopart A, Long M: **The origin of the Jingwei gene and the complex modular structure of its parental gene, yellow emperor, in Drosophila melanogaster.** *Mol Biol Evol* 2000, **17**:1294-1301.
44. Long M, Langley CH: **Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila.** *Science* 1993, **260**:91-95.
45. Schurko AM, Mazur DJ, Logsdon JMJ: **Inventory and phylogenomic distribution of meiotic genes in Nasonia vitripennis and among diverse arthropods.** *Insect Mol Biol* 2010, **19**(Suppl 1):165-180.
46. Fernandez D, Valdivia A, Irazusta J, Ochoa C, Casis L: **Peptidase activities in human semen.** *Peptides* 2002, **23**:461-468.
47. Schaller J, Glander HJ: **Flow cytometric analysis of enzymes in live spermatozoa before and after cryostorage.** *Andrologia* 2000, **32**:357-364.
48. Yasuhara T, Yokosawa H, Ishii S: **Purification and characterization of an aminopeptidase from sperm of the sea urchin, Strongylocentrotus intermedius. Ca2(+)-dependent substrate specificity as a novel feature of the enzyme.** *J Biochem (Tokyo)* 1990, **107**:273-279.
49. Togo T, Morisawa M: **GPI-anchored aminopeptidase is involved in the acrosome reaction in sperm of the mussel Mytilus edulis.** *Mol Reprod Dev* 2004, **67**:465-471.
50. Vrbancovski MD, Lopes HF, Karr TL, Long M: **Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes.** *PLoS Genet* 2009, **5**:e1000731.
51. Jao SC, Huang LF, Hwang SM, Li WS: **Tyrosine 387 and arginine 404 are critical in the hydrolytic mechanism of Escherichia coli aminopeptidase P.** *Biochemistry* 2006, **45**:1547-1553.
52. Strater N, Lipscomb WN: **Transition state analogue L-leucinephosphonic acid bound to bovine lens leucine aminopeptidase: X-ray structure at 1.65 Å resolution in a new crystal form.** *Biochemistry* 1995, **34**:9200-9210.

53. Piatigorsky J: **Gene Sharing and Evolution**. Cambridge: Harvard University; 2007.
54. Matsui M, Fowler JH, Walling LL: **Leucine aminopeptidases: diversity in structure and function**. *Biol Chem* 2006, **387**:1535-1544.
55. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**:1596-1599.
56. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572-1574.
57. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions**. *Mol Biol Evol* 1986, **3**:418-426.
58. Pond SL, Frost SD, Muse SV: **HyPhy: hypothesis testing using phylogenies**. *Bioinformatics* 2005, **21**:676-679.
59. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search**. *Anal Chem* 2002, **74**:5383-5392.

doi:10.1186/1471-2164-12-177

Cite this article as: Dorus *et al.*: Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of *Drosophila* sperm. *BMC Genomics* 2011 **12**:177.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

